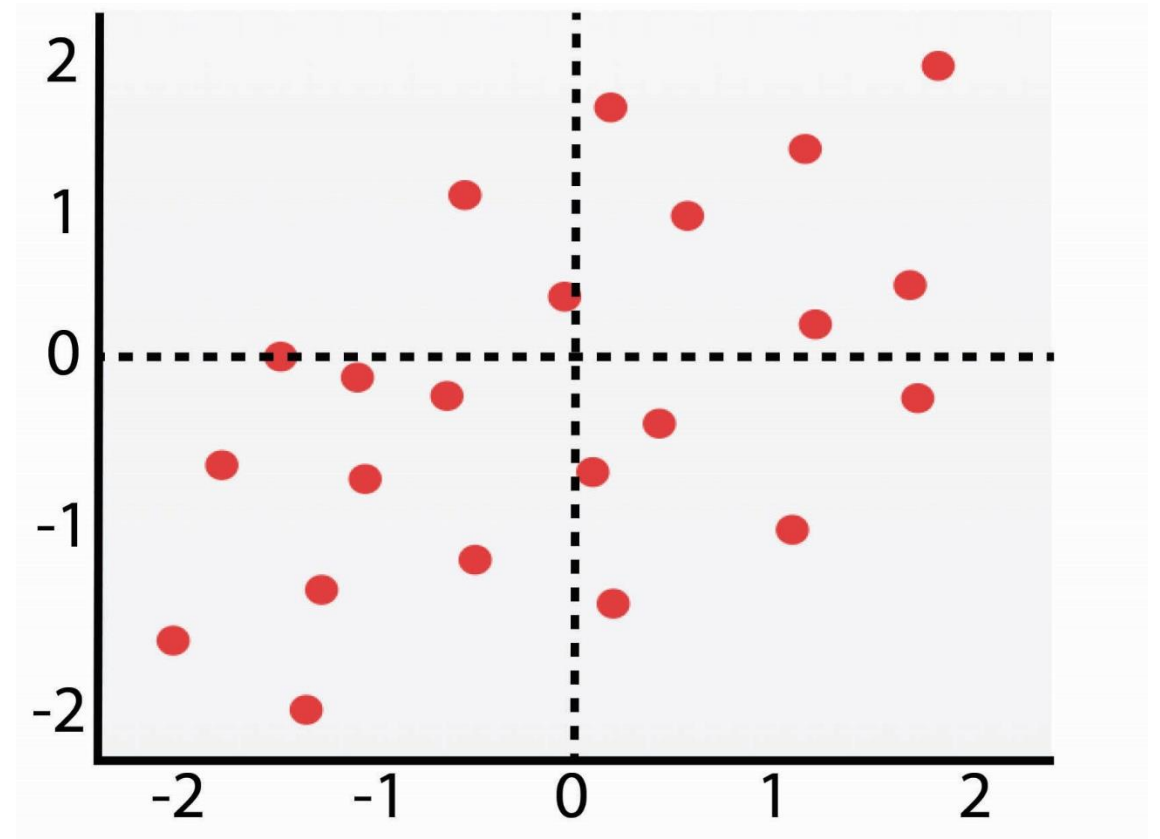


Correlation

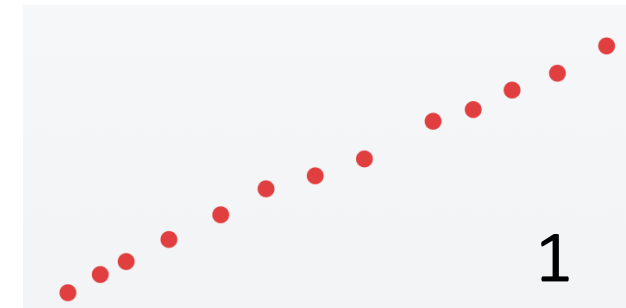
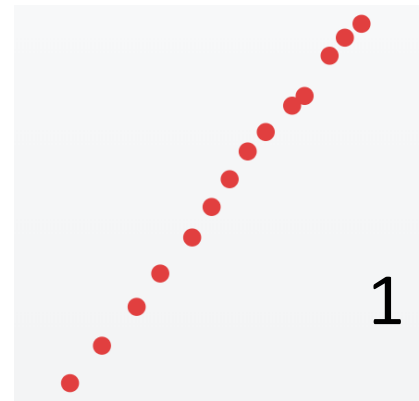
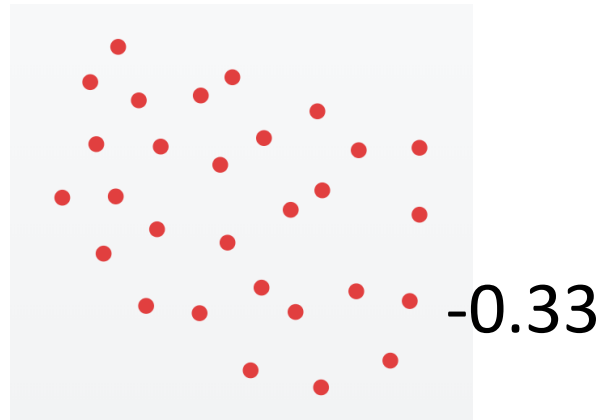
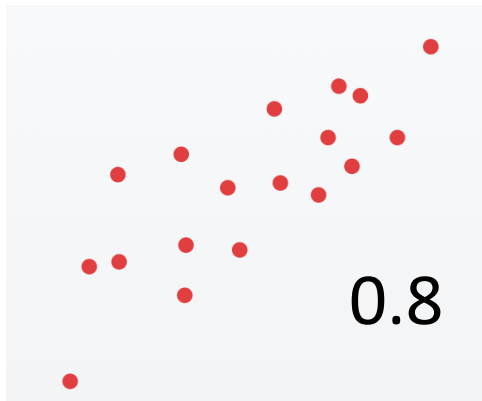
The **correlation** between two quantitative variables measures the strength and direction of their linear relationship, expressed by the **correlation coefficient** r .

$$r = \frac{\Sigma(z_x z_y)}{n - 1}$$

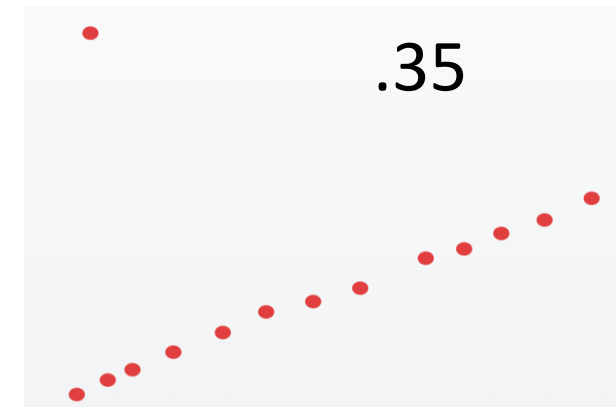


Facts about the correlation coefficient r

1. If r is positive, the scatterplot goes up from left to right. If r is negative, the scatterplot goes down from left to right.
2. r is always between -1 (perfect negative correlation) and 1 (perfect positive correlation). $r=0$ means no correlation.



3. Switching the variables has no effect on r .
4. Shifting or stretching the data has no effect on r .
5. Correlation has no units.
6. Correlation is sensitive to outliers.



7. Correlation does not imply causation!

8. Non-linear relationships can have strange correlations. Four possibilities:

a) Scatterplot is linear and tightly packed: high correlation

b) Scatterplot is linear and cloud-shaped: low correlation

c) Scatterplot is curved and tightly packed: high association, but we need to calculate r differently.

d) Scatterplot is curved and cloud-shaped: low association, but we need to calculate r differently.

